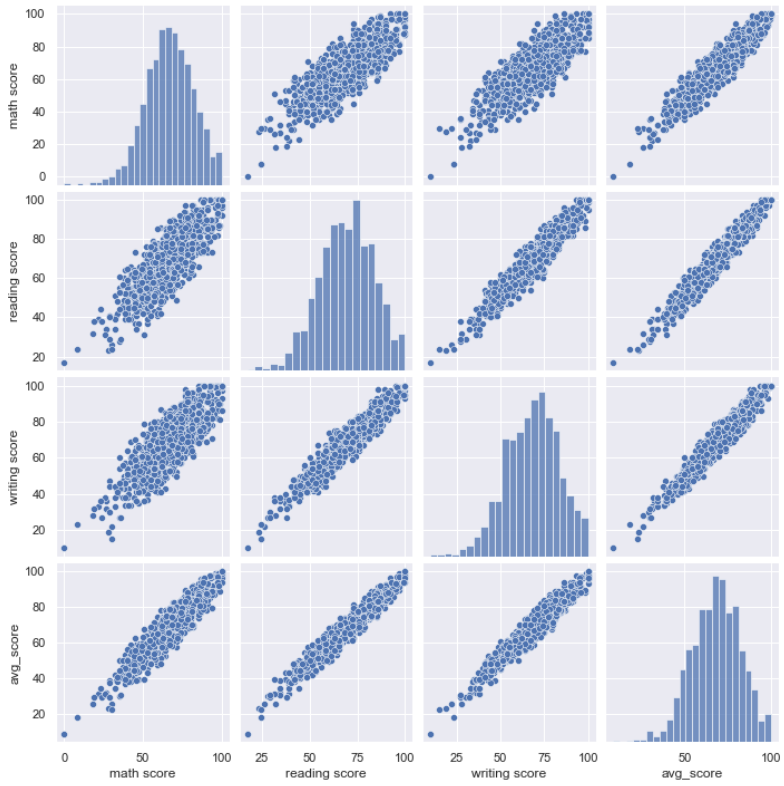
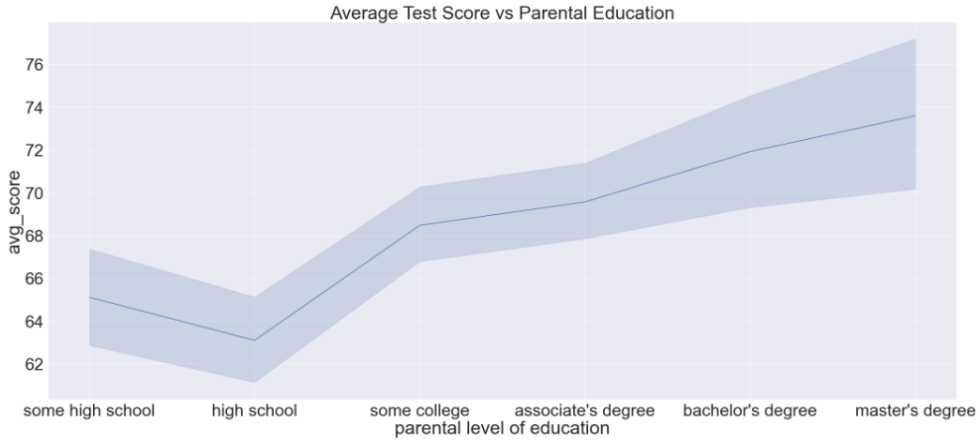
My dataset is from Kaggle and contains data on one thousand students' family education history, lunch type(standard or free/reduced), demographics, and study habits along with their scores on math, reading, and writing tests. I want to explore the most important factors of test scores. This can give us information on what the most important factors to academic success are and help improve the effectiveness of social programs by showing what areas of life should be targeted.

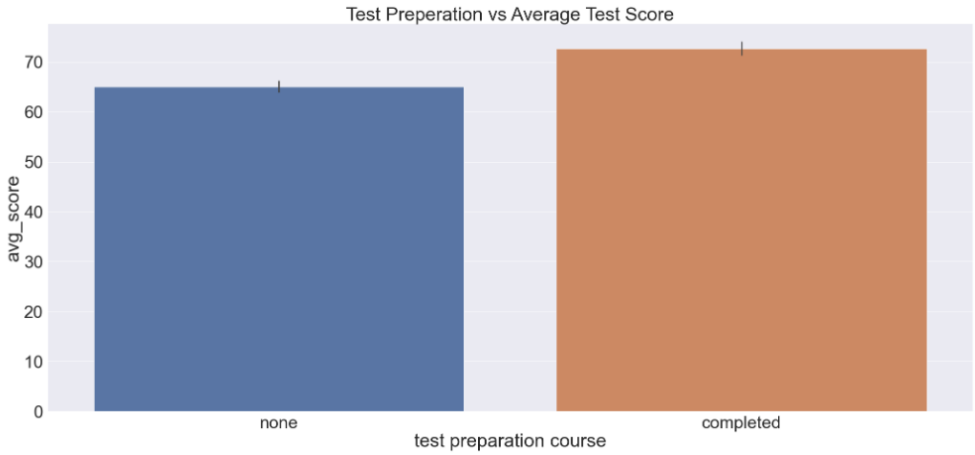


My preprocessing consisted of exploring how each variable was measured. Instead of listing individual's race/ethnicity, they put everyone into 5 groups(A,B,C,D,E). The test preparation course variable is just a "completed" or "none" value. While a continuous value would've been more useful for me, I think it can still be useful. Each test score is out of 100 points. The variable measuring parents' level of education is the most interesting to me. It ranges anywhere from "some high school" to "master's degree". The one visualization I show in the preprocessing section shows how well each test score correlates with every other one. Because all three correlate closely, I combined them into one variable called "avg_score". Making one visualization instead of three will save a lot of time and space. The last part of preprocessing is setting the parental level of education variable in order from least amount of school to greatest amount of school. If there is a clear trend between an increase in parental education and an increase in test scores, it will be much easier to see.

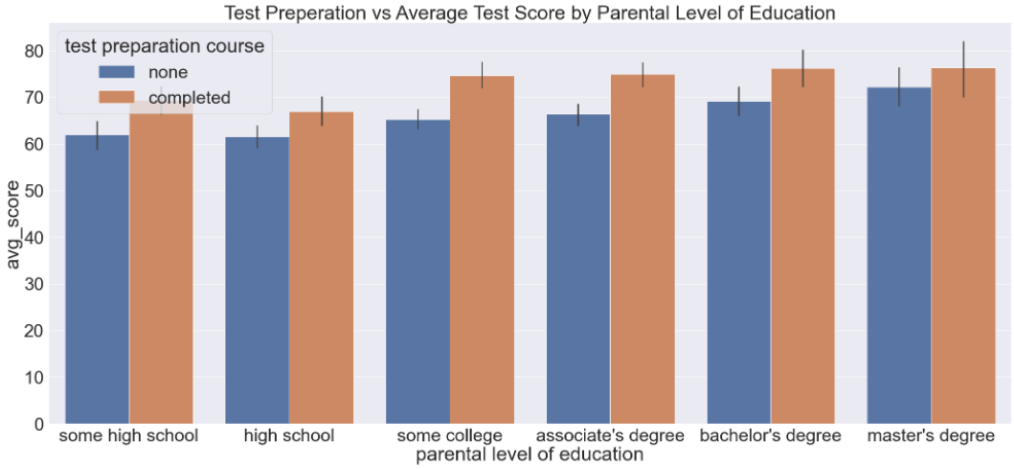
The first visualization uses the parental education variable to see how the parental education correlates with test scores. The figure below shows a very clear trend between the two variables. As the amount of education that someone's parents have, the higher their average test score will be for every single test.



I wasn't sure how much someone's parents directly had to do with them doing well on a test, so I looked at how someone prepared for a test compared to how well they did on it. The figure to the right shows that visualization. Both Test Preparation and Parental Education level have similar trends when correlation with test scores so exploring them at the same time along with test scores is the next logical step.



Using the visualization below, we can see that as Parental education increases, overall test scores go up. Additionally, we see that within each category, test scores go up when the test preparation is completed.



What we can draw from this is that growing up in a household

that is educated is better for increasing test scores. This could be for a multitude of reasons, people that are educated will likely pass the value of education down to their children, leading to higher test scores. That could also lead to a higher proportion of kids from more educated families preparing for the tests more, raising the average score. Access is a very important part of this conversation. Having parents that value education and are around more often to help further their kids' education outside of the classroom also tend to already be more highly educated. Giving all kids access to educational resources and mentoring outside of school could be a potential solution, or at least part of the solution to this problem. No one deserves to be at a disadvantage because of their parents' education.